Evaluation of Machine Learning Techniques for Security in SDN

Ahnaf Ahmad*, Erkki Harjula*, Mika Ylianttila*, Ijaz Ahmad†
* University of Oulu, Oulu, Finland
†VTT Technical Research Center of Finland, Espoo, Finland
Email: *ahnaf15ahmad@gmail.com, *[firstname.lastname]@oulu.fi,†[firstname.lastname]@vtt.fi

2020 IEEE Globecom Workshops (GC Wkshps) | 978-1-7281-7307-8/20/$31.00 ©2020 IEEE | DOI: 10.1109/GCWkshps50303.2020.9367477

***Abstract*—Software Defined Networking (SDN) has emerged as the most viable programmable network architecture to solve many challenges in legacy networks. SDN separates the network control plane from the data forwarding plane and logically centralizes the network control plane. The logically centralized control improves network management through global visibility of the network state. However, centralized control opens doors to security challenges. The SDN control platforms became the most attractive venues for Denial of Service (DoS) and Distributed DoS (DDoS) attacks. Due to the success and inevitable benefits of Machine Learning (ML) in fingerprinting security vulnerabilities, this article proposes and evaluates ML techniques to counter DoS and DDoS attacks in SDN. The ML techniques are evaluated in a practical setup where the SDN controller is exposed to DDoS attacks to draw important conclusions for ML-based security of future communication networks.**

***Index Terms*—SDN; Security; IDS; DDoS; Machine Learning; Security in SDN**

I. Introduction

The growth of Internet has ushered in a new era of connectivity, control, and security. Due to the rapid progress in new services such as e-health, e-commerce, and unmanned aerial vehicles, etc., that require sophisticated network policies and complex networking tasks, the need for a new and better way of managing communication networks has developed over time [1]. Most of the existing challenges of communication networks, such as static nature and complexity in management, have been addressed with the introduction of Software Defined Networking (SDN). SDN allows to manage the network in a much simpler way and the concept provides the network system more software-based control rather than hardware-based. It separates the control and data planes and provides more flexibility in the management of the network system.

The SDN architecture has three main architectural planes, termed as the application plane, the control plane, and the data plane. The application plane is responsible for creating policies, network management rules, and Quality of Service (QoS) for the controller. The Control plane is responsible for traffic engineering, traffic management, network management, and so on. It is the most important plane of the SDN architecture. The data plane consists of elements that form an underlying network to forward network traffic. The application and control planes are connected through north-bound interfaces, whereas the control and data planes are connected through south-bound interfaces [2].

Although SDN introduced simpler and more convenient networks, it also introduced new security vulnerabilities. These vulnerabilities can lead to security threats that can be catastrophic for the SDN's architecture in particular, and the whole network in general [3], as shown in Fig. 1. The security of SDN is crucial to future networks. As SDN centralizes the control of the entire network through logically centralized control platforms, working of the entire network is dependent on those controllers. Albeit the benefits, the separation of the planes makes it easy to fingerprint the controllers and target it for Denial of Service (DoS) or Distributed DoS (DDoS) attacks [4]. DDoS is one of the major attacks in the network system. It is growing constantly with new ways to attack the system. More hosts are vulnerable to these attacks as the Internet is surging ahead.

Intrusion Detection System (IDS) overviews the traffic, analyses it, and detects anomalies or unauthorized access in the network domain [5]. Several features are used for attack detection in SDN. Deviation from normal traffic flow and high rate of traffic are the two important features to detect DDoS attacks. Normally, IDS constantly analyses the network traffic and uses a lot of resources while doing that. With the emergence of 5G,

978-1-7281-7307-8/20/$31.00 ©2020 IEEE

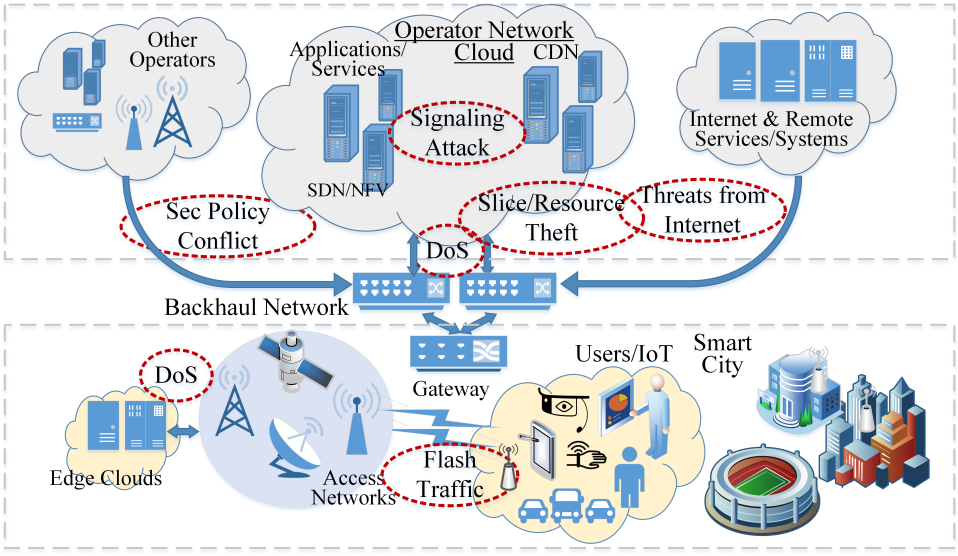

Fig. 1. Potential security challenges in 5G.

the response time and accuracy to these attacks are quite significant and the network system needs to be updated constantly to overcome these challenges. The centralized network architecture of SDN offers the opportunity to create more efficient solutions [6].

IDS based on Machine Learning (ML) has been on the rise for the past few years. ML-based IDS creates a pattern from several features of the traffic dataset and learns to differentiate attack from the normal traffic. ML-based IDS is more accurate and efficient compared to traditional approaches. In this article an evaluation of ML techniques in IDS is carried out using several ML algorithms, a unique dataset and a module that works with the centralized controller. ML techniques such as Support Vector Machine (SVM), Naive-Bayes, Decision Tree, and Logistic Regression are used in the experiments. The evaluation of different models provides a brief idea about the efficiency of the algorithms, implementation strategy, and how to improve it.

This article is organized as follows: In section II, the background of the work is discussed with security in SDN, the use of ML for IDS in SDN, and previous related work. Section III discusses the ML techniques, the experimental setup of the environment, and evaluation of the results. The article is concluded in Section IV.

II. Background

In the era of Internet of Things (IoT), security has become a very important factor in network systems [7]. Cybersecurity plays a huge role in today's network whether it is for a company or a single user. Intrusion is the attempt to compromise a network system's confidentiality, integrity, and availability. It can be even an attempt to bypass the network security mechanism. A network intrusion is a form of active attack in the network system. Intrusion in the network system has grown exponentially over time. The process of discovering unauthorized access within the network is known as Intrusion Detection. A system that monitors the traffic and detects malicious activity in the network is called Intrusion Detection System (IDS) [8].

A. SDN Security

The logically centralized control of the SDN provides a new dimension of opportunities in the network system. It increases the performance of a network system as well as its programmability. The platform has also brought new security challenges [6]. There are several issues that concern the SDN system.

- **Forwarding Device Attack:** This attack can be done through access points and switches with attacks like DoS and it can cause failure in network system or disrupt it for a certain period of time.
- **Control Plane Threats:** As SDN is logically centralized, attack on the control plane can put down the entire network system. Control plane threats are one of the biggest issues in SDN security.
- **Communication Vulnerabilities:** The communication channel in the OpenFlow protocol has its own security protocol such as TLS for data-control, but it

can be disabled by the administration. This creates the opportunity for man-in-the-middle attacks.

- **Fake Traffic Flows:** It is a big challenge as an attacker can create fake traffic to overflow the network channel and create disruption. DoS and DDoS attacks are used to overload network resources and disrupt the system.
- **Open Programmable APIs:** Open Programmable APIs gave the platform a big advantage as well as creating new challenges regarding network security. This issue must be managed with certain protocols.

The SDN architecture has its unique requirements of security and lacking in these requirements make the architecture vulnerable to various attacks and threats [9]. Therefore, it is very important to take necessary steps to make sure that all the components of SDN are secured. As the SDN controller manages the whole network system, it is necessary to ensure that the controller is safe. Otherwise, it can cause the failure of the whole network system. Dynamically updated security measures are crucial to keep the network system secure from different threats and attacks.

B. Machine Learning for SDN Security

ML has become a very important technology in the telecommunication field [10]. The implementation of ML in SDN has been a notable aspect of the platform. ML is used to attain key structural patterns and models from the training data of the system. It mainly consists of two phases. The training phase uses ML methods to learn the pattern and creates the model from the dataset for detecting network anomalies. The later phase is decision-making where the model predicts and acts according to the input training data. Intrusion detection is a classification task in the ML-based IDS. Supervised ML is mostly applied in IDS. Variation in dimensions of the input, such as flow features, impacts the performance of the ML algorithms. The global view gives the opportunity to learn the traffic flow easily in the network as well as react quickly to an attack.

C. Related Work

DDoS attacks are one of the major threats in IDS. The aim is to exhaust the network resources and lead to the unavailability of the network system. The research for intrusion with DDoS attacks has been ongoing for a long time and implementation of ML has been an important step. Several methods have been tried and tested to detect DDoS attacks in SDN.

In [11], SVM, Naive Bayes, J48, and Random Forest ML methods were benchmarked with public dataset. While comparing, J48 showed the highest accuracy which was 80%. The author also mentioned the importance of feature selection and dataset labelling. Author in [12] used C4.5 (Decision Tree), Bayesian Network, Naive Bayes, and Decision Table for benchmarking. Historical network data was used for predicting the attack and Bayesian Network had an accuracy of 91.68%. The Longtail project 19 data was used for this project.

In [13], the observation of Multilayer perception, SVM, Decision Tree, and Random Forest was done with flow table attack, bandwidth attack, and controller attack. Overall, Decision Tree was better but Random Forest had better accuracy. The author used 11 different features in which only 5 were critical. In [14], the author used 6 features and labelled the traffic in the dataset for training. SVM, Naive Bayes, and k-Nearest Neighbours (KNN) have been used for the experiment while KNN had an accuracy of 97%.

In [15], the author evaluated different ML algorithms based on feature selection. SVM, KNN, Artificial Neural Network (ANN), and Naive Bayes were used. The paper focused on the accuracy of the methods with different feature selection. The use of feature selection method improved the accuracy very little. Precision and F1 score varied quite a bit. The authors were also aware of having the balance between resource usage and accuracy because higher the feature set, higher the usage of resources.

The accuracy of the ML-based IDS is better. They also perform well with better resource management. Algorithm selection is important. Researchers are working with different algorithms evaluating their performances. Determination of features for the IDS depends a lot on the algorithm used. ML algorithms learn the pattern of the features of the traffic flow and classifies the traffic accordingly. This gives the ML-based IDS an advantage over the non-ML based IDS. Non-ML based IDS focuses on the deviation of behaviour of the traffic flow. Sometimes, they can detect high bandwidth usage of the network as abnormal traffic.

III. Evaluation of ML Techniques

ML algorithms are thoroughly used in data mining applications that allow users to learn about patterns and models from the data. It is thoroughly used in anomaly-based IDS by training the model to detect intrusion in the system. There are four different ML techniques which are supervised, unsupervised, semi-supervised,

and reinforcement learning [16]. Supervised learning method is commonly used in anomaly-based IDS and intrusion detection is considered as a classification task. Input of the dataset for the training of the model has an impact on accuracy. The resource management is an important factor for IDS. Increasing the number of features in dataset increases the accuracy but at the same time uses more resources from the system. This slows down the process. It is important to keep the balance between performance and resource management.

A. Selected Methods

ML-based methods use different techniques to detect anomaly in the system. Various types of network features are taken into account while classifying the traffic. The detection is based on that behavioral learning pattern. Based on the different approaches of learning patterns of the ML algorithms, SVM, Naive Bayes, Decision Tree, and Logistic Regression are chosen for the experiment.

- **Support Vector Machine (SVM):** SVM is a discriminative classifier which is defined by separating a hyperplane. The objective is to find a suitable hyperplane that can distinguish between the data points [17]. Hyperplanes are decision boundaries that allow classifying the data points. A separate kernel function is used to perform mapping of the data. SVM can learn with very limited amount of data and provide good results.
- **Naive-Bayes:** The algorithm is based on Bayes' Theorem. It assumes the predictors to be independent among themselves [18]. The classifier assumes that the presence of a feature in a class is unrelated to other features. The amount of data has an impact on the algorithm. Differentiation is done by creating maps of each class value through a list of instances that belongs to the class. Conditional probability is used to predict an attack and normal traffic.
- **Decision Tree:** This algorithm takes decision by learning simple decision rules from the training data [19]. It uses the tree structure for classification. The whole dataset is divided into small subsets. The main objective of the algorithm is to pursue the best classification rate. The decision tree algorithm builds small trees for better resource management.
- **Logistic Regression:** The classification is done by assigning observation to a discrete set of classes. Logistic sigmoid function is used to recur a probability that maps two or more discrete classes. The prediction analysis is based on the concept of probability [20].

B. Experimental Setup

The experiment is based on a simulation environment. The experimental setup consists of an HP laptop with an Intel® Core™ i7-8500U CPU @1.80GHz and 1.99GHz processor with 16GB of RAM and 512GB of Solid State Drive. A virtual machine with Ubuntu 18.04 is used. A remote controller with a tree-based topology having a depth of two is used to create the testbed with Mininet. POX controller and Open vSwitches with OpenFlow protocol are used.

The dataset is created with the traffic flow generated by Scapy. UDP flood attack is used for the DDoS attack alongside normal traffic. The dataset is marked with normal and attack traffic. Several hosts were used to initiate a normal traffic flow as well as attack flow to analyse the detection capability of good and bad traffic. The controller was able to detect and differentiate between good and bad traffic. Single host as well as multi host attacks were used in the network system.

C. Evaluation of Results

The traffic flow during the experiment is shown in Fig. 2. The high peaks are attacks and the low peaks are normal traffic flows from single and multiple hosts.

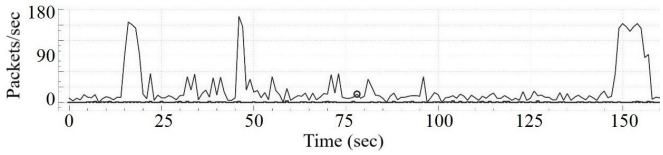

Fig. 2. Flow of packets.

The analysis of different ML approaches can be seen in Fig. 3 with the percentage of accuracy, precision, and error.

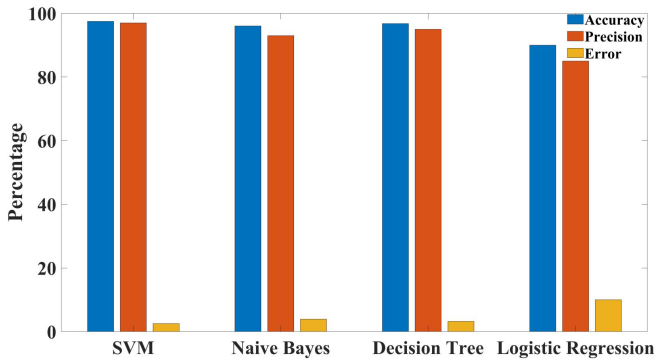

Fig. 3. Performance analysis of ML techniques.

TABLE I
PERFORMANCE OF DIFFERENT ML MODELS

ML Algorithms	Accuracy	Sensitivity	Specificity	Precision	F1-Score
SVM	97.50%	98.27%	96.70%	97 %	96%
Naive-Bayes	96.03%	95.95%	92.63%	93 %	94%
Decision Tree	96.78%	97.11%	94.62%	95 %	95%
Logistic Regression	89.98%	91.62%	84.57%	85 %	86%

The detailed analysis with sensitivity, specificity, and F1-Score is in Table I, along with accuracy and precision. The result shows that the SVM performed better than other algorithms. It has an accuracy of 97.5% and a precision of 97% with a very minimum error rate of 2.5%. All the algorithms have an accuracy of over 96% except Logistic Regression.

Receiver Operating Characteristics (ROC) curve is a tool that can be used to evaluate the test results. Fig. 4 is a two-dimensional graph with the True Positive (TP) rate on the Y-axis and the False Positive (FP) rate on the X-axis. The ROC curve is a trade-off between sensitivity (TP rate) and specificity (1 - FP rate). The algorithm having an ROC curve closer to the top left corner indicates better performance.

In Table II, the Area Under Curve (AUC) for different algorithms is given. A good model has an AUC close to 1 which means that it has a good measure of separability. All four models have an AUC of over 0.9. Thus, the models have good distinguishability between the two classes of attack and normal traffic. The ROC curve of all 4 ML methods is shown in Fig. 4. From the ROC

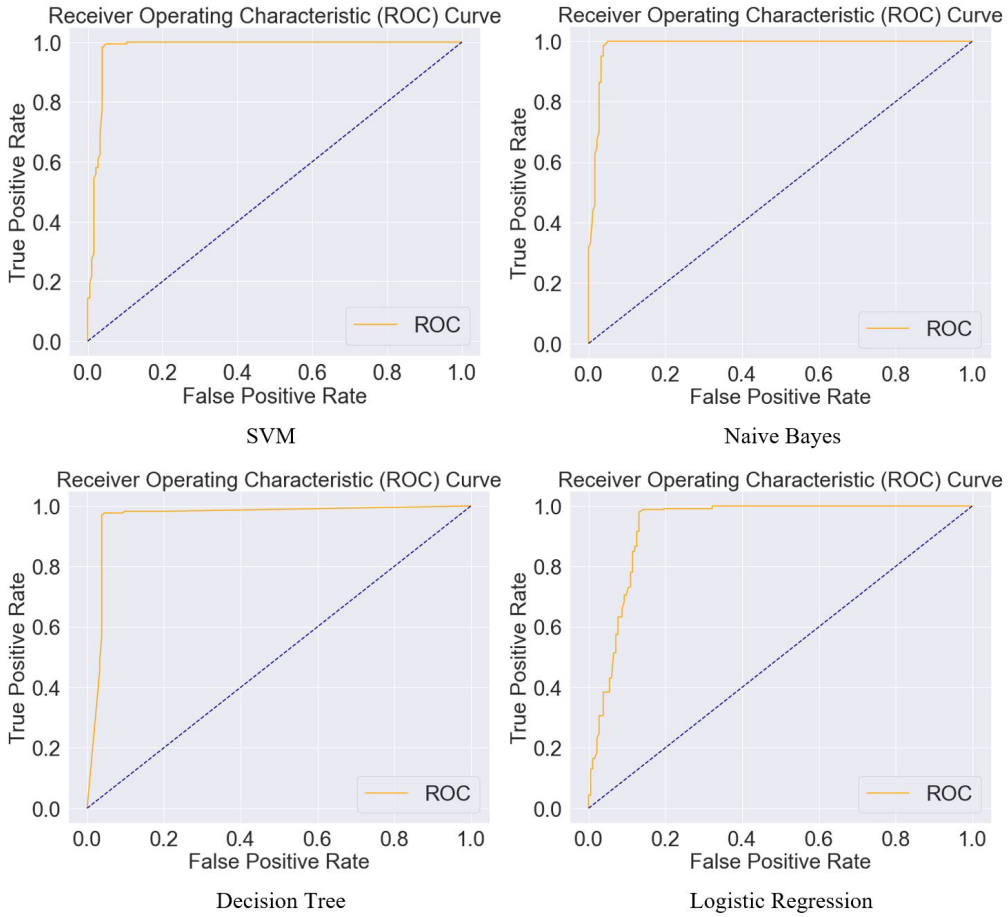

Fig. 4. ROC curve of different ML algorithms.

curve, the SVM, Naive Bayes, and Decision Tree show better performance than the Logistic Regression. The range of AUC for the ROC curve between 0.9 and 1 is considered excellent. In Table II, the AUC for the ROC of all the methods are well over 0.9 and Naive Bayes has the highest value of 0.99.

TABLE II
AREA UNDER CURVE (AUC) OF ML ALGORITHMS

ML Algorithms	AUC
SVM	0.98
Naive-Bayes	0.99
Decision Tree	0.96
Logistic Regression	0.93

IV. CONCLUSION

SDN facilitates the implementation of ML for IDS in several ways. Most of the existing implementations are carried out with public datasets that are not updated regularly. The dataset of this experiment was created from the simulation environment. Three out of four algorithms have over 96% accuracy while SVM attained 97.5%. It also did not mix up high bandwidth use with a DDoS attack. Important future work includes evaluation of scalability of the control platform when different ML techniques are deployed in the control platform for DDoS detection and intrusion prevention along IDS.

ACKNOWLEDGMENT

This work has been supported by TEKES Finland and Academy of Finland under projects: 6Genesis Flagship (grant 318927), and 5GEAR (Grant No. 319669) projects, and SecureConnect.

REFERENCES

[1] I. Ahmad, T. Kumar, M. Liyanage, M. Ylianttila, T. Koskela, T. Braysy, A. Anttonen, V. Pentikinen, J.-P. Soininen, and J. Huusko, "Towards gadget-free internet services: A roadmap of the naked world," *Telematics and Informatics*, vol. 35, no. 1, pp. 82 – 92, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0736585316305597

[2] I. Ahmad, S. Namal, M. Ylianttila, and A. Gurtov, "Towards software defined cognitive networking," in *2015 7th International Conference on New Technologies, Mobility and Security (NTMS)*, 2015, pp. 1–5.

[3] M. Liyanage, I. Ahmed, J. Okwuibe, M. Ylianttila, H. Kabir, J. L. Santos, R. Kantola, O. L. Perez, M. U. Itzazelaia, and E. M. De Oca, "Enhancing security of software defined mobile networks," *IEEE Access*, vol. 5, pp. 9422–9438, 2017.

[4] I. Ahmad, T. Kumar, M. Liyanage, J. Okwuibe, M. Ylianttila, and A. Gurtov, "Overview of 5G Security Challenges and Solutions," *IEEE Communications Standards Magazine*, vol. 2, no. 1, pp. 36–43, MARCH 2018.

[5] H.-J. Liao, C.-H. R. Lin, Y.-C. Lin, and K.-Y. Tung, "Intrusion detection system: A comprehensive review," *Journal of Network and Computer Applications*, vol. 36, no. 1, pp. 16–24, 2013.

[6] I. Ahmad and S. Namal and M. Ylianttila and A. Gurtov, "Security in Software Defined Networks: A Survey," *IEEE Communications Surveys Tutorials*, vol. 17, no. 4, pp. 2317–2346, Fourthquarter 2015.

[7] I. Ahmad, S. Shahabuddin, T. Kumar, J. Okwuibe, A. Gurtov, and M. Ylianttila, "Security for 5G and Beyond," *IEEE Communications Surveys Tutorials*, vol. 21, no. 4, pp. 3682–3722, Fourthquarter 2019.

[8] R. A. Kemmerer and G. Vigna, "Intrusion detection: a brief history and overview," *Computer*, vol. 35, no. 4, pp. supl27–supl30, 2002.

[9] S. Namal, I. Ahmad, A. Gurtov, and M. Ylianttila, "Enabling Secure Mobility with OpenFlow," in *2013 IEEE SDN for Future Networks and Services (SDN4FNS)*, 2013, pp. 1–5.

[10] I. Ahmad, S. Shahabuddin, T. Kumar, E. Harjula, M. Meisel, M. Juntti, T. Sauter, and M. Ylianttila, "Challenges of AI in Wireless Networks for IoT," *arXiv preprint arXiv:2007.04705*, 2020.

[11] M. S. Elsayed, N. Le-Khac, S. Dev, and A. D. Jurcut, "Machine-learning techniques for detecting attacks in sdn," in *2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT)*, 2019, pp. 277–281.

[12] S. Nanda, F. Zafari, C. DeCusatis, E. Wedaa, and B. Yang, "Predicting network attack patterns in sdn using machine learning approach," *2016 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*, pp. 167–172, 2016.

[13] R. Santos, D. Silva, W. Santo, A. Ribeiro, and E. Ordonez, "Machine learning algorithms to detect ddos attacks in sdn," *Concurrency and Computation: Practice and Experience*, p. e5402, 06 2019.

[14] A. Prakash and R. Priyadarshini, "An intelligent software defined network controller for preventing distributed denial of service attack," in *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, 2018, pp. 585–589.

[15] H. Polat, O. Polat, and A. Çetin, "Detecting ddos attacks in software-defined networks through feature selection methods and machine learning models," *Sustainability*, vol. 12, p. 1035, 02 2020.

[16] J. Xie, F. R. Yu, T. Huang, R. Xie, J. Liu, C. Wang, and Y. Liu, "A survey of machine learning techniques applied to software defined networking (sdn): Research issues and challenges," *IEEE Communications Surveys Tutorials*, vol. 21, no. 1, pp. 393–430, 2019.

[17] T. Evgeniou and M. Pontil, "Support vector machines: Theory and applications," in *Advanced Course on Artificial Intelligence*. Springer, 1999, pp. 249–257.

[18] H. Zhang, "The optimality of naive bayes," vol. 2, 01 2004.

[19] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE transactions on systems, man, and cybernetics*, vol. 21, no. 3, pp. 660–674, 1991.

[20] J. Peng, K. Lee, and G. Ingersoll, "An introduction to logistic regression analysis and reporting," *Journal of Educational Research - J EDUC RES*, vol. 96, pp. 3–14, 09 2002.